1 Introduction

The post-war history of income inequality in advanced countries can be divided, at least roughly, into two phases. From 1945 to about the mid-1980s, pre-tax inequality decreased at least in part because of a reduction in skilled/unskilled wage differentials and asset inequality. The second phase occurred from the 1980s onwards, when inequality reversed course and increased. Using the Luxembourg Income Study (LIS) data, Immervoll and Richardson (2011) reported that in Organisation for Economic Co-operation and Development (OECD) countries, government redistribution has become less effective in compensating for increasing inequalities since the 1990s (see also Figure 1). Moreover, top income shares have increased in many advanced economies over the past three decades (see Figure 2),and top tax rates on upper-income earners have declined significantly in many OECD countries during this period (see Figure 3).

How can we explain this evolution of redistribution in OECD countries? For example, Atkinson, Piketty and Saez (2011) emphasised that it is very difficult to account for the rise in top incomes using the standard labour supply/demand explanation. Hence, the role of social policies and progressive taxation should not be dismissed in these discussions. In fact, some of the basic features of redistribution can be explained through the optimal tax framework developed by Mirrlees (1971). This model has dominated the economics of redistributive taxation for the past forty years. It captures the central features in thinking about the development of redistribution policy. Three elements of the model are useful for this purpose. First is the concept of inherent inequality. If there is no intervention by the government, inherent inequality will be fully reflected in disposable income. However, if the government wants to intervene – as seems to be the case in developed countries – we will find the second component of the Mirrlees model, the egalitarian objectives of the government. Moreover, if the government tries to redistribute income from high-income people to low-income people, there will be incentive and disincentive effects. In other words, the redistribution policy is the product of circumstances and objectives.

The recent optimal income tax literature has put a lot of emphasis on top marginal tax rates. As is well known, optimal income tax literature provides a striking result on a top marginal tax rate. The optimal marginal tax rate for the highest-wage person is zero. This result – due to Sadka (1976) – actually says that the highest income should be subject to a zero marginal tax rate. Strictly speaking, this result applies only to a single person at the very top of the income distribution, suggesting it is a mere theoretical curiosity. Moreover, it is unclear

Cambridge University Press
978-1-108-46911-1 — Inequality and Optimal Redistribution
Hannu Tanninen , Matti Tuomala , Elina Tuominen
Excerpt
More Information

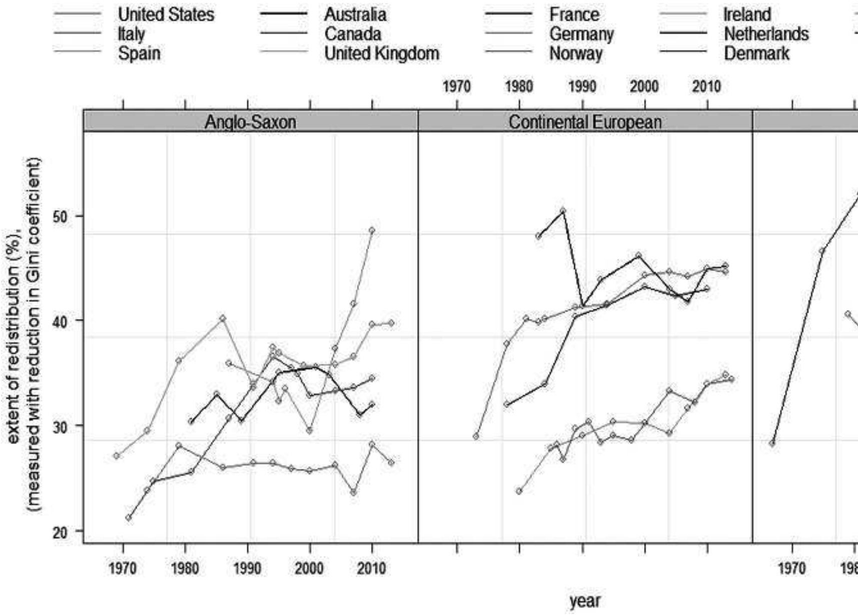

Figure 1 Evolution of the extent of redistribution measured with the relative reduction in the
countries (unbalanced data over the years 1967–2013). Authors' calculations fr

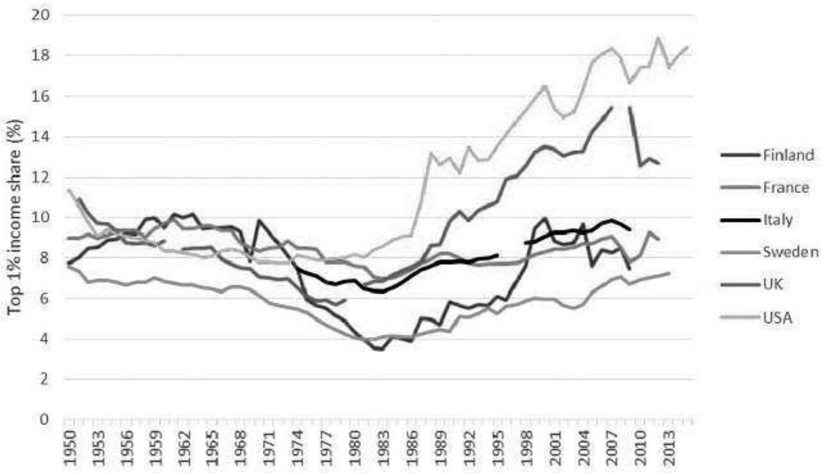

Figure 2 Evolution of the top 1 per cent pre-tax income shares in six countries. Data source: The World Inequality Database (formerly the World Wealth and Income Database, http://wid.world, accessed 1 April 2016)

that a ‘top earner’ even exists. For example, Saez (2001) argues that ‘unbounded distributions are of much more interest than bounded distributions to address the high income optimal tax rate problem.’ Without a top earner, the intuition for the zero top marginal rate does not apply, and marginal rates near the top of the income distribution may be positive and even large. Moreover, calculations in Tuomala (1984) show that the zero rate is not a good approximation for high incomes.

Notably, almost all analytical results focus on the structure of marginal tax rates to the neglect of average tax rates, which are arguably more important indicators of income tax progressivity.[1] Computational techniques can be utilised to say something about average rates. Moreover, the optimal tax structure is usually depicted in terms of skills rather than incomes, which is more relevant for actual policy recommendation. The derivation of optimal tax rates based on income is difficult, because the income of the tax function that is maximised is an endogenous variable depending on the tax function itself, so indirect effects have to be taken into account.

Starting around 1980, almost all developed countries have seen a sharp decline in tax progressivity. At the same time, many developed countries abolished annual or inherited wealth taxes. Moreover, a growing fraction of capital income was gradually left out of the income tax base. Consequently, only

[1] There are some exceptions; see Boadway and Jacquet (2008).

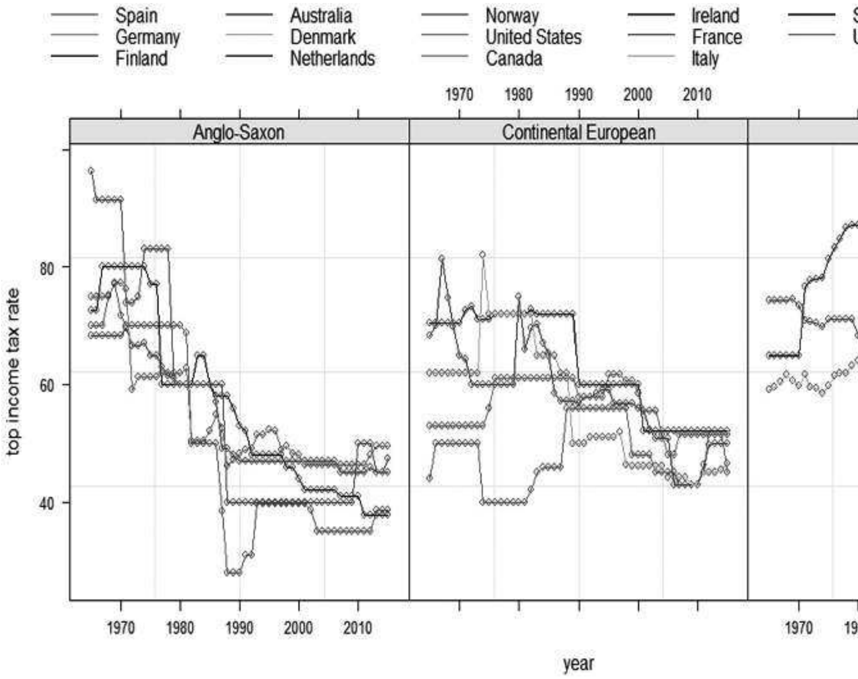

Figure 3 Evolution of top income tax rates in fourteen advanced countries over the years 1965–
Stantcheva (2014), OECD (2017) and the Association of Finnish Local and Regi

a labour income tax is any more progressive in the tax system. In the Nordic dual income tax system, this has been done explicitly. Dual income tax systems in turn have suffered from income shifting from progressively taxed labour income to capital income, which is taxed at a lower, flat rate.[2] We have also seen a rising share of capital in many advanced countries since the 1980s (see Piketty and Zucman, 2014). This, in turn, has increased overall income inequality because ownership of capital is much more unequally distributed than labour income. Therefore, equity considerations suggest capital income should be taxed more than labour income. If capital accumulation is sensitive to the net-of-tax return, these considerations in turn suggest going in a different direction. Moreover, capital is more mobile internationally than labour. Given these considerations, how should capital income be taxed? Roughly classifying, we can distinguish four alternatives to taxing capital income: not at all, linearly (Nordic dual income tax), relating the marginal tax rates of capital and labour incomes and taxing all income on the same schedule.

We survey some of the earlier results in linear and nonlinear taxation and produce some new numerical results both in the standard Mirrlees model and in its extensions. We consider how the optimal income tax schedule changes when income inequality, and in particular top income inequality, increases. We focus our attention both on the top tax rate (often thought of as the top marginal rates) and on the entire tax schedule. In particular, we consider how the optimal average tax schedule changes when top income inequality increases. Much of the discussion of optimal income taxation is about labour income. However, given the key role of capital income in overall income inequality, we also consider the optimal taxation of capital income, and in particular the taxation of top capital income.

We also examine empirically the relationship between the extent of redistribution and the elements of the Mirrlees model, including market income inequality, by utilising the LIS database and the government's redistributive preferences. The LIS database provides data on both factor and disposable incomes for a number of advanced countries over the past three to four decades, which facilitates the study of the extent of redistribution. In our empirical specifications, we study two alternative indicators to measure income inequality, namely the Gini coefficient and the percentile ratio (P90/P50). To measure the extent of redistribution, we use the relative reduction between the inequality indicator for factor incomes and disposable incomes. Previously, Tanninen and Tuomala (2005) examined the relationship between inherent inequality and the

[2] In Norway, for example, there is a progressive element in the capital income tax since shareholders pay a tax on what is defined as above-normal return on shares. Other capital income is taxed linearly. The Finnish dual income tax system is not purely linear anymore.

extent of redistribution by utilising the LIS data for a number of OECD countries over two to three decades. They found that redistribution in these countries is positively associated with inherent inequality until the mid-1990s. However, their empirical results were based on the assumption that the degree of espoused egalitarianism has remained constant over the period considered. There is now some recent individual-country evidence that there could have been a shift in norms, causing governments to become less willing to finance transfers and to levy progressive taxes, leading to reductions in the extent of redistribution. One could argue, in line with Atkinson (1999), that these kinds of changes have been episodic rather than time-trend, and are therefore rather difficult to justify, for example, in the context of median voter models. Thus, we focus here also on the role of the egalitarian objectives of government, which is an important component of the optimal income tax model. We construct our redistributive preference measure using the optimal top tax formula (for given labour supply elasticity) for which we have collected data from various sources.

The remainder of this Element is organised as follows. Section 2 analyses optimal linear income with different social objectives when inequality varies. Section 3 sets up the basic Mirrlees (1971) model and highlights the role of different elements of the model in determining the optimal redistribution. Section 4 analyses top marginal tax rates in the case of quasi-linear preferences, Pareto-distributed skills and constant labour supply elasticities. We also study how elements left out of the standard model change top tax rates. Using numerical simulations, we study in Section 5 the role of different social objectives when inherent (or pre-tax inequality) income inequality increases. Section 6 analyses separable taxes on labour income and capital income in the simplified framework. Moreover, we briefly outline the case of optimal comprehensive income tax. Section 7 examines empirically the relationship between the extent of redistribution and the components of the optimal nonlinear tax model. In Section 8, we extend nonlinear taxation with the Veblen effect and analyse briefly the redistributive role of factors such as publicly provided private goods (health, education, social services), public employment, endogenous wages in the overlapping generations (OLG) model and income uncertainty that are missing in the standard model. Section 9 concludes.

2 Optimal Linear Labour Income Taxation

We start by studying optimal linear taxes. In the linear income tax system, the tax is characterised by a lump sum income or a basic income B paid to each

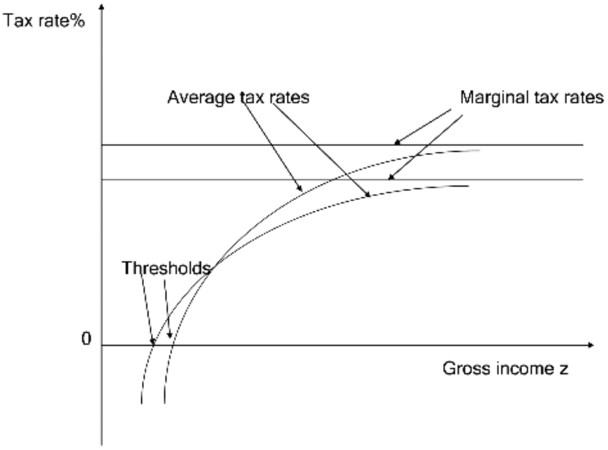

Figure 4 Marginal and average tax rates and linear income taxation

individual and a proportional tax on each euro earned at a rate t (the flat rate). If $t > 0$ and $B > 0$, the linear income tax is progressive in the sense that the average tax rate rises over the entire income range (see Figure 4). The linear income tax schedule provides a minimum guaranteed income to individuals whose income falls short of the critical level. This is the feature of the linear income tax system which leads us to refer to the section of the tax schedule below some gross income level z* in Figure 4 as a negative income tax.[3] The negative income tax system, basic income or the social dividend system has been proposed, supported and widely discussed by several distinguished economists such as James Meade, Milton Friedman, James Tobin and Tony Atkinson.[4] In particular, in a developing country context, linear income taxation can be justified as an easily implementable instrument.

Individuals face a linear income tax schedule $T(z) = -B + tz$. Every individual in this model faces a budget constraint $x(n) = (1 - t)z(n) + B$, where $(1 - t)$ is the net of tax rate, l is labour supply (l could be interpreted as the number of hours worked by the individual or equally well as effort), $z = nl$ is before-tax income. The revenue requirement of the government, R, to be used for expenditure on public goods, is taken as given. The government's budget constraint is:

3 There is an interesting simpler extension of the linear income tax: a two-bracket income tax. It applies a constant rate t_1 to all income up to some specified level z* and another constant rate t_2 to all income over the specified level z*. See Slemrod et al. (1994), Apps et al. (2013).

4 For some history on negative income tax and related proposals, see Kesselman and Garfinkel (1978).

$$\int_0^\infty [-B + tnl(1-t, B)]f(n)dn = R, \tag{1}$$

where $Z = \int_0^\infty nlf(n)dn$ is the aggregate labour supply or income supply from the population.

The central issue considered in the analysis of the optimal linear income tax is that of choosing between the basic income and the associated tax rate. Therefore it is plausible to express individuals' preferences in terms of the indirect utility function, denoted $V(n(1-t), B)$. In making this choice, the government is assumed to be constrained by a government budget and by the responses of taxpayers. The taxpayers are assumed to adjust their labour supply in response to changes in taxation.

In a typical optimal tax analysis, the objective for policy is to maximise social welfare, an object that is calculated through a social welfare function (SWF) that depends only upon the welfare levels of the individuals in society. The government has redistributive objectives represented by $W(V^1, \ldots, V^N)$ with $W' > 0$, $W'' < 0$. This is called prioritarianism by philosopher Derek Parfit (1991) (see also Matthew Adler, 2012). The idea of prioritarianism is that just distributions require giving greater weight, or priority, to individuals who are worse off. In fact, optimal tax theory has long made use of prioritarianism. In the founding paper of the modern optimal tax literature Mirrlees (1971) specified an objective for tax policy that directly translated the core idea of prioritarianism into his model's formal mathematical language.

The government's problem is to choose B and t so as to maximise the SWF:

$$\int_0^\infty W(V(n(1-t), B))f(n)dn \tag{2}$$

under the budget constraint (1).

From the first-order condition of this problem we arrive at the condition (see the derivation in Tuomala (1985, 2016)

$$\frac{t}{1-t} = \frac{1}{E}[1 - \frac{z(\psi)}{\bar{z}}], \tag{3}$$

where $E = \frac{d\bar{z}(1-t)}{d(1-t)\bar{z}}$ (the elasticity of earnings with respect $1-t$, net-of-tax and averaging over the taxpayers must give $\bar{z} = \bar{z}(1-t, B)$), $\psi = W_V V_B$ is the social marginal utility of income and $z(\psi) = \int_0^\infty \psi z f(n)dn / \int_0^\infty \psi f(n)dn$.

To illustrate the formula in (3) further, we have to specify the key elements of the model. We concentrate on the special case where there are no income effects on labour supply and the elasticity of labour supply with respect to the net-of-tax wage rate is constant. If ε denotes this elasticity, the quasi-linear indirect utility function is given by:

$$V(n(1-t),B) = B + \frac{[n(1-t)]^{1+\varepsilon}}{1+\varepsilon} \tag{4}$$

To simplify, we assume that the social marginal valuation depends only on wage (ability or productivity) n and not on the level of utility. We adopt a constant relative inequality aversion form of the welfare function. It is also called the Atkinson social welfare function by Adler (2012). The contribution to social welfare of the individual is $\frac{n^{1-\gamma}}{1-\gamma}$, where γ is the constant relative inequality aversion coefficient. In other words, the SWF is a quasi-concave function of n. Hence the social marginal value of income to an individual with wage rate n is proportional to $n^{-\gamma}$. It gives us the utilitarian case, where $\gamma = 0$, since we are back with the sum. Utilitarianism gives no value to equality in the distribution of well-being. It cares only about the total of well-being, not about how well-being is spread amongst the people. Rawls' formulation of the objective may be seen as a limiting case of the iso-elastic function as γ tends to infinity. Hence W takes the form min u, i.e. maximin. Rawls (1974) objects to this interpretation. For him, it is wrong to suggest that we can 'shift smoothly from the moral conception to another simply by varying the parameter' (γ). (Rawls, 1974, p. 664). Rawls (1974) suggests that the important feature of a distributive criterion is that it should serve as a public principle. He says that 'citizens generally should be able to understand it and have some confidence that it is realized' (Rawls, 1974, p. 143). He claims that the maximin, unlike utilitarianism, satisfies this criterion of sharpness or transparency. Hence, a change in tax policy that benefits the least advantaged should be easily observable.

The key axiomatic difference between utilitarianism and prioritarianism is the *Pigou-Dalton axiom* (axiom of transfers), here understood in terms of well-being. With $\gamma = 0$, the SWF is no longer prioritarian. Pigou-Dalton is not satisfied.[5]

Next we turn to the distribution of n. The excellent Pareto fit of the top tail of the distribution has been well known for more than a century, since the pioneering work of Pareto (1896), and has been verified in many countries and many

[5] Pigou-Dalton: A gap-diminishing transfer of well-being from someone better off to someone worse off, leaving everyone else unaffected, is an ethical improvement.

periods, as summarised in Atkinson et al. (2011). In those twenty-four countries reported in Atkinson et al. (2011), the Pareto parameter typically varies between 3.0 and 1.67. The top tail of the income distribution is closely approximated by a Pareto distribution.[6] The higher α (i.e. lower coefficient α/(α-1); i.e. less fat upper tail) implies lower inequality. A lower coefficient means larger top income shares and higher income inequality. In Finland during the period 1990–2014, the Pareto parameter (taxable income) varied between 3.7 (1992) and 1.79 (2004) (see Figure 13 in Section 6).

We assume here that the n-distribution is an unbounded Pareto distribution $f(n) = \frac{1}{n^{1+a}}$ for a > 0, i.e. a Pareto tail with the coefficient α. Thus, the right tail is thicker as α is smaller, implying that only low-order moments exist. The Pareto parameter in itself is an appropriate measure for increasing top income shares. Using the property of the Pareto distribution $\mathrm{E}(n^j) = \frac{\alpha n_0^j}{\alpha - j}$, we can calculate the values of the optimal tax rate and of the basic income from the following formula:

$$\frac{t}{1-t} = \frac{1}{\varepsilon}[1 - M] \tag{5}$$

where $M = \left\{ \left[\frac{1-\frac{1+\varepsilon}{\alpha}}{1-\frac{1+\varepsilon}{\alpha+\gamma}} \right] \right\}$

Substituting the labour supply function $l = [n(1-t)]^{\varepsilon}$ for the revenue constraint, we can express the basic income and the revenue relative to the average earnings (denoted by b and r, respectively). We rewrite the revenue constrains as follows: $b = t(1-t)^{\varepsilon} - r$. The results are presented in Table 1. The revenue requirement is set to zero, thus the system is purely redistributive. Results are shown for two different values of labour supply elasticity and for two different values regarding income dispersion, $\alpha = 2$ and $\alpha = 2.5$. The tax rates are high for all the combinations of parameter values.

Most work on optimal nonlinear and linear income taxation used the lognormal distribution to describe the distribution of productivities $ln(n; m, \sigma^2)$ with support [0,∞) with parameters m and σ (see Aitchison and Brown, 1957). The first parameter, m, is log of the median and the second parameter is the variance of log wage. The latter one is itself an inequality measure. Using the property of the lognormal distribution $\ln E(n^j) = jm + j^2\sigma^2/2$, we can obtain the optimal tax rate formula $\frac{t}{1-t} = \frac{1}{\varepsilon}[1 - e^{-\gamma(1+\varepsilon)\sigma^2}]$ or using the property of lognormal

6 It is still the case that the original purpose of the Pareto function is its most fruitful application. This view is nicely expressed by Cowell (1977) when he writes that: 'Although the Pareto formulation has proved to be extremely versatile in the social sciences, in my view the purpose for which it was originally employed is still its most useful application – an approximate description of the distribution of income and wealth among the rich and the moderately rich.'